

Министерство науки и высшего образования РФ
Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

СОГЛАСОВАНО

Заведующий кафедрой

**Кафедра прикладной
математики и компьютерной
безопасности (ПМКБ_ИКИТ)**

наименование кафедры

подпись, инициалы, фамилия

«___» _____ 20__ г.

институт, реализующий ОП ВО

УТВЕРЖДАЮ

Заведующий кафедрой

**Кафедра прикладной математики
и компьютерной безопасности
(ПМКБ_ИКИТ)**

наименование кафедры

А.А. Кытманов

подпись, инициалы, фамилия

«___» _____ 20__ г.

институт, реализующий дисциплину

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
БАЗОВЫЕ МОДЕЛИ И МЕТОДЫ
МАШИННОГО ОБУЧЕНИЯ (THE
BASICS OF MACHINE LEARNING)**

Дисциплина Б1.В.02 Базовые модели и методы машинного обучения
(The Basics of Machine Learning)

Направление подготовки / 01.04.02 Прикладная математика и
специальность информатика, программа 01.04.02.09 Data
Science and Mathematical Modeling 2020г

Направленность
(профиль)

Форма обучения

очная

Год набора

2020

Красноярск 2021

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

составлена в соответствии с Федеральным государственным образовательным стандартом высшего образования с учетом профессиональных стандартов по укрупненной группе

010000 «МАТЕМАТИКА И МЕХАНИКА»

Направление подготовки /специальность (профиль/специализация)

Направление 01.04.02 Прикладная математика и информатика,
программа 01.04.02.09 Data Science and Mathematical Modeling 2020г.

Программу
составили

1 Цели и задачи изучения дисциплины

1.1 Цель преподавания дисциплины

Целью изучения дисциплины является ознакомление студентов с основными методами и моделями анализа данных.

1.2 Задачи изучения дисциплины

Задачами изучения дисциплины являются:

- Ознакомление с задачами обучения на размеченных данных, основными линейными моделями, методами оценки качества прогнозирования, возможными причинами низкого качества моделей;
- Выработка умений постановки задач анализа данных, обоснованного выбора и применения изученных методов в прикладных исследованиях
- Получение практических навыков программирования алгоритмов машинного обучения средствами языка R.

1.3 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

УК-2:Способен управлять проектом на всех этапах его жизненного цикла.
УК-2.1:Формулирует на основе поставленной проблемы проектную задачу и способ ее решения через реализацию проектного управления.
УК-2.2:Разрабатывает концепцию проекта в рамках обозначенной проблемы: формулирует цель, задачи, обосновывает актуальность, значимость, ожидаемые результаты и возможные сферы их применения.
УК-2.3:Планирует необходимые ресурсы, в том числе с учетом их заменимости.
УК-2.4:Разрабатывает план реализации проекта.
УК-2.5:Осуществляет мониторинг хода реализации проекта, корректирует отклонения, вносит дополнительные изменения в план реализации проекта, уточняет зоны ответственности участников проекта.
УК-1:Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий.
УК-1.2:Определяет пробелы в информации, необходимой для решения проблемной ситуации, и проектирует процессы по их устранению.
УК-1.3:Критически оценивает надежность источников информации, работает с противоречивой информацией из разных источников.
ПК-4:Способен разрабатывать и внедрять новые методы и технологии исследования данных.
ПК-4.1:Знает: состояние и перспективы развития информационных технологий, технологий данных в России и в мире; современные и перспективные методы сбора, хранения и передачи данных; источники данных, интенсивность генерации данных источниками; технические средства и среды сбора, хранения

и обработки данных; современные и перспективные средства визуализации и интерпретации данных; исследование операций; машинное обучение; математическое моделирование; методы сравнительного анализа.

1.4 Место дисциплины (модуля) в структуре образовательной программы

Продвинутые методы анализа данных

Анализ больших данных

Прикладные задачи анализа данных

Статистическое моделирование

выполнение и защита выпускной квалификационной работы

Продвинутые методы анализа данных (Advanced Methods of Data Analysis)

Технологическая (проектно-технологическая) практика (Field Internship II)

Анализ больших данных (Big Data)

Машинное обучение и криптография (Applications of Machine Learning in Cryptography)

Прикладные задачи анализа данных (Applied Data Analysis)

Статистическое моделирование (Statistical Modeling)

1.5 Особенности реализации дисциплины

Язык реализации дисциплины Английский.

Дисциплина (модуль) реализуется без применения ЭО и ДОТ.

2. Объем дисциплины (модуля)

Вид учебной работы	Всего, зачетных единиц (акад.час)	Семестр
		1
Общая трудоемкость дисциплины	6 (216)	6 (216)
Контактная работа с преподавателем:	2 (72)	2 (72)
занятия лекционного типа	1 (36)	1 (36)
занятия семинарского типа		
в том числе: семинары		
практические занятия	1 (36)	1 (36)
практикумы		
лабораторные работы		
другие виды контактной работы		
в том числе: групповые консультации		
индивидуальные консультации		
иная внеаудиторная контактная работа:		
групповые занятия		
индивидуальные занятия		
Самостоятельная работа обучающихся:	3 (108)	3 (108)
изучение теоретического курса (ТО)		
расчетно-графические задания, задачи (РГЗ)		
реферат, эссе (Р)		
курсовое проектирование (КП)	Нет	Нет
курсовая работа (КР)	Нет	Нет
Промежуточная аттестация (Экзамен)	1 (36)	1 (36)

3 Содержание дисциплины (модуля)

3.1 Разделы дисциплины и виды занятий (тематический план занятий)

№ п/п	Модули, темы (разделы) дисциплины	Занятия лекционного типа (акад. час)	Занятия семинарского типа		Самостоятельная работа, (акад. час)	Формируемые компетенции
			Семинары и/или Практические занятия (акад. час)	Лабораторные работы и/или Практикумы (акад. час)		
1	2	3	4	5	6	7
1	Basic components of machine learning	4	4	0	14	
2	Linear methods of regression	8	8	0	22	
3	Linear methods of classification	10	8	0	24	
4	Variance and Covariance analysis	6	6	0	22	
5	Model quality issues	8	10	0	26	
Всего		36	36	0	108	

3.2 Занятия лекционного типа

№ п/п	№ раздела дисциплины	Наименование занятий	Объем в акад. часах		
			Всего	в том числе, в инновационной форме	в том числе, в электронной форме
1	1	The ingredients of machine learning	2	0	0
2	1	Data Preparation	2	0	0
3	2	Probabilistic justification of the regression model	2	0	0
4	2	Multiple linear regression model	4	0	0
5	2	Metrics in regression models	2	0	0

6	3	Classification problem. Metrics for classification	2	0	0
7	3	Bayesian classifier. Linear discriminant analysis. Logistic Regression.	4	0	0
8	3	Support vector classifier. Support vector machine	2	0	0
9	3	Linear perceptron	2	0	0
10	4	Analysis of variance	2	0	0
11	4	Analysis of covariance	4	0	0
12	5	Multicollinearity problem	2	0	0
13	5	Feature selection. Principal component analysis	2	0	0
14	5	Overfitting problem. Regularization.	2	0	0
15	5	Outliers and missed values	2	0	0
Всего			26	0	0

3.3 Занятия семинарского типа

№ п/п	№ раздела дисципли ны	Наименование занятий	Объем в акад. часах		
			Всего	в том числе, в инновационной форме	в том числе, в электронной форме
1	1	Basics of programming using R	2	0	0
2	1	Descriptive statistics and visual data analysis	2	0	0
3	2	Multiple linear regression with Least square method	4	0	0
4	2	Quantile regression. Metrics of regression	4	0	0
5	3	Linear discriminant analysis. Logistic Regression.	4	0	0
6	3	Support vector classifier. Support vector machine	2	0	0
7	3	Elemental perceptron	2	0	0
8	4	Analysis of covariance	6	0	0
9	5	Feature selection for classification model with Caret package	2	0	0
10	5	Stepwise feature selection in regression models	4	0	0

11	5	Principle component analysis	2	0	0
12	5	Outlier and high leverage points detection	2	0	0
Всего			26	0	0

3.4 Лабораторные занятия

№ п/п	№ раздела дисциплины	Наименование занятий	Объем в акад. часах		
			Всего	в том числе, в инновационной форме	в том числе, в электронной форме
Всего					

4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

	Авторы, составители	Заглавие	Издательство, год
Л1.1	Рубан А. И.	Методы анализа данных: учебное пособие	Красноярск: СФУ, 2012
Л1.2	Зарова Е. В.	Applied Multivariate Statistical Analysis: Presentations for Lecturing and Working Examples with R=Прикладной многомерный статистический анализ: Презентации для лекций и примеры решений с использованием пакета R: Учебное пособие на английском языке	Москва: ООО "Научно-издательский центр ИНФРА-М", 2016

5 Фонд оценочных средств для проведения промежуточной аттестации

Оценочные средства находятся в приложении к рабочим программам дисциплин.

6 Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

6.1. Основная литература			
	Авторы, составители	Заглавие	Издательство, год
Л1.1	Neter J., Wasserman W., Kutner M. H.	Applied Linear Statistical Models: Regression, Analysis of Variance and Experimental Designs	Boston: Irwin, 1990
Л1.2	Dunn P. F.	Measurement and Data Analysis for Engineering and Science: научное издание	Boca Raton: CRC Press, 2010
6.2. Дополнительная литература			

	Авторы, составители	Заглавие	Издательство, год
Л2.1	Кобзарь А. И.	Прикладная математическая статистика. Для инженеров и научных работников	Москва: Физматлит, 2006
Л2.2	Кабаков Р., Волкова П. А.	R в действии. Анализ и визуализация данных на языке R	Москва: ДМК Пресс, 2014
Л2.3	Джеймс Г.	Введение в статистическое обучение с примерами на языке R	Москва: ДМК Пресс, 2017
Л2.4	Волкова П. А., Шипунов А. Б.	Статистическая обработка данных в учебно-исследовательских работах: Учебное пособие	Москва: Издательство "ФОРУМ", 2016
6.3. Методические разработки			
	Авторы, составители	Заглавие	Издательство, год
Л3.1	Рубан А. И.	Методы анализа данных: учебное пособие	Красноярск: СФУ, 2012
Л3.2	Зарова Е. В.	Applied Multivariate Statistical Analysis: Presentations for Lecturing and Working Examples with R=Прикладной многомерный статистический анализ: Презентации для лекций и примеры решений с использованием пакета R: Учебное пособие на английском языке	Москва: ООО "Научно- издательский центр ИНФРА- М", 2016

7 Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля)

Э1	1. Professional informational and analytical resource dedicated to machine learning, pattern recognition and data mining Machinelearning.ru. Режим доступа: http://www.machinelearning.ru	http://www.machinelearning.ru
----	---	---

8 Методические указания для обучающихся по освоению дисциплины (модуля)

Для успешного усвоения дисциплины и допуска к экзамену необходимо:

- Прослушать все лекции или, в случае пропуска занятий, изучить из материал самостоятельно по рекомендуемым источникам литературы;
- Выполнить все 9 индивидуальных заданий и защитить преподавателю в форме собеседования;
- Выполнить групповой проект и защитить его в форме докладов на семинаре.

Экзамен сдается в устной форме. Каждый экзаменационный билет состоит из 2-х теоретических вопросов из списка вопросов к экзамену.

9 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю) (при необходимости)

9.1 Перечень необходимого программного обеспечения

9.1.1	Для выполнения индивидуальных заданий по программированию алгоритмов анализа данных необходима программная среда вычислений R версии не младше, чем 3.4.3, а также редактор кода RStudio версии 1.1 или старше.
-------	---

9.2 Перечень необходимых информационных справочных систем

9.2.1	1. Система электронного обучения Сибирского федерального университета (https://e.sfu-kras.ru)
9.2.2	2. электронные информационно-справочные ресурсы научной библиотеки СФУ (http://bik.sfu-kras.ru)
9.2.3	3. Интернет-репозиторий статистических данных Data-Planet (https://www.data-planet.com/)
9.2.4	4. Интернет-репозиторий статистических данных на платформе kaggle.com. Режим доступа: https://www.kaggle.com/datasets

10 Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине (модулю)

Учебные лаборатории и классы, оснащенные современными компьютерами, объединенными в локальные вычислительные сети с выходом в Интернет, а также периферийным и проекционным оборудованием.